

**BIOSURVEILLANCE UNTUK DETEKSI PENYEBARAN  
PENYAKIT DI INDONESIA BERDASARKAN TWEETS  
MENGUNAKAN ALGORITMA NAÏVE BAYES**

**TUGAS AKHIR**

**Sebagai Persyaratan Guna Meraih Gelar Sarjana Strata 1  
Teknik Informatika Universitas Muhammadiyah Malang**



**Oleh :**

**LAZQAR MARKUS OKTAVIANTO**

**201310370311189**

**JURUSAN TEKNIK INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS MUHAMMADIYAH MALANG  
2018**

**LEMBAR PERSETUJUAN**

**BIOSURVEILLANCE UNTUK DETEKSI PENYEBARAN  
PENYAKIT DI INDONESIA BERDASARKAN TWEETS  
MENGUNAKAN ALGORITMA NAÏVE BAYES**

**LAZQAR MARKUS OKTAVIANTO**

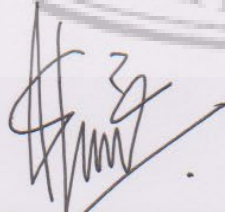
**(201310370311189)**

Sebagai Persyaratan Guna Meraih Gelar Sarjana Strata 1  
Teknik Informatika Universitas Muhammadiyah Malang

Malang, 29 November 2018

Menyetujui,

Dosen I



Setio Basuki, S.T., M.T.  
NIDN : 0714028403

Dosen II



Yufis Azhar, S.Kom., M.Kom.  
NIDN : 0728088701

## LEMBAR PENGESAHAN

### BIOSURVEILLANCE UNTUK DETEKSI PENYEBARAN PENYAKIT DI INDONESIA BERDASARKAN TWEETS MENGUNAKAN ALGORITMA NAÏVE BAYES

#### TUGAS AKHIR

Sebagai Persyaratan Guna meraih Gelar Sarjana Strata 1  
Teknik Informatika Universitas Muhammadiyah Malang

Disusun Oleh :

**LAZQAR MARKUS OKTAVIANTO**

**(201310370311189)**

Tugas Akhir ini telah diuji dan dinyatakan lulus melalui sidang majelis penguji  
pada tanggal 27 Oktober 2018

Menyetujui,

Penguji I,

Penguji II,

Gita Indah Marthasari, S.T., M.Kom.

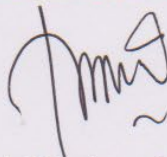
Fauzi Dwi S. S., S.T., M.CompSc.

NIP : 108.0611.0442

NIDN: 0707069202

Mengetahui,

Ketua Jurusan Teknik Informatika



Gita Indah Marthasari, S.T., M.Kom.

NIP: 108.0611.0442

## LEMBAR PERNYATAAN

Yang bertanda tangan dibawah ini :

**NAMA : LAZQAR MARKUS OKTAVIANTO**

**NIM : 201310370311189**

**FAK./JUR. : TEKNIK / TEKNIK INFORMATIKA**

Dengan ini saya menyatakan bahwa Tugas Akhir dengan judul **BIOSURVEILLANCE UNTUK DETEKSI PENYEBARAN PENYAKIT DI INDONESIA BERDASARKAN TWEETS MENGGUNAKAN ALGORITMA NAÏVE BAYES** beserta seluruh isinya adalah karya saya sendiri bukan merupakan karya tulis orang lain, baik sebagian maupun seluruhnya, kecuali dalam bentuk kutipan yang telah disebutkan sumbernya.

Demikian surat pernyataan ini saya buat dengan sebenar-benarnya. Apabila kemudian ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya ini, atau ada klaim dari pihak lain terhadap keaslian karya saya ini maka saya siap menanggung segala bentuk resiko/sanksi yang berlaku.

Mengetahui,

Malang, 5 Desember 2018

Dosen Pembimbing

Yang Membuat Pernyataan

**Setio Basuki, S.T., M.T.**

**NIDN : 0714028403**



**Lazqar Markus Oktavianto**

## KATA PENGANTAR

Alhamdulillah, puji syukur kehadiran Allah SWT, yang telah melimpahkan rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan tugas akhir yang menjadi salah satu syarat untuk menyelesaikan program studi Teknik Informatika jenjang Strata-1 Universitas Muhammadiyah Malang (UMM). Sholawat serta salam semoga tetap tercurah kepada Nabi besar Muhammad saw, keluarga, sahabat dan para pengikutnya hingga akhir jaman.

Penulis menyadari bahwa dalam menyelesaikan tugas akhir ini tidak lepas dari peran berbagai pihak yang telah banyak memberi bantuan, nasehat, bimbingan dan dukungan. Dalam kesempatan ini penulis ingin mengucapkan terima kasih yang tak terhingga khususnya kepada :

1. Orang Tua Penulis atas segala do'a restu dan dukungannya baik material atau spiritual kepada penulis dalam penyelesaian tugas akhir ini.
2. Bapak Setio Basuki, S.T., M.T. dan Bapak Yufis Azhar, S.Kom., M.Kom. selaku Dosen Pembimbing I dan II tugas akhir. Terima kasih atas pengarahan yang telah di berikan sehingga dapat menyelesaikan laporan tugas akhir ini dengan benar adanya.
3. Teman-teman yang membantu dan selalu memberikan dukungan juga meluangkan waktunya dalam proses penulisan tugas akhir ini.

Penulis menyadari bahwa dalam penyusunan tugas akhir ini banyak kekurangannya. Oleh karena itu, penulis mengharapkan kritik dan saran yang dapat menyempurnakan penulisan ini sehingga dapat bermanfaat untuk pengembangan ilmu.

Malang, 5 Desember 2018

Penulis

## DAFTAR ISI

LEMBAR PERSETUJUAN .....	i
LEMBAR PENGESAHAN .....	ii
LEMBAR PERNYATAAN .....	iii
KATA PENGANTAR.....	iv
ABSTRAK .....	v
ABSTRACT .....	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR.....	x
DAFTAR TABEL .....	xi
DAFTAR GRAFIK.....	xii
BAB I.....	1
PENDAHULUAN .....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah .....	2
1.3 Tujuan.....	3
1.4 Batasan Masalah.....	3
1.5 Metode Penelitian.....	4
1.5.1 Studi Pustaka.....	4
1.5.2 Analisis Kebutuhan .....	4
1.5.3 Desain Sistem.....	4
1.5.4 Implementasi.....	4
1.5.5 Pengujian dan Evaluasi .....	4
1.5.6 Pembuatan Laporan .....	5
1.6 Sistematika Penulisan.....	5

<b>BAB II .....</b>	<b>6</b>
<b>LANDASAN TEORI .....</b>	<b>6</b>
2.1 Penelitian Terdahulu.....	6
2.2 Biosurveillance .....	7
2.3 Text Mining .....	7
2.4 Twitter .....	8
2.5 Pre-processing Data.....	9
2.6 Fitur Sintaktik.....	10
2.7 Classification .....	10
2.8 Algoritma Naïve Bayes Classifier .....	11
2.9 Named Entity Recognition .....	12
2.10 INA-NLP .....	12
2.11 Metode Pengujian.....	13
2.11.1 Cross Validation.....	13
2.11.2 Confusion Matrix .....	13
<b>BAB III.....</b>	<b>15</b>
<b>ANALISIS DAN PERANCANGAN SISTEM.....</b>	<b>15</b>
3.1 Biosurveillance .....	15
3.2 Data Penelitian.....	15
3.3 Analisis Data .....	16
3.4 Preprocessing Data .....	17
3.5 Perancangan Pelatihan Klasifikasi NBC .....	17
3.6 Mendapatkan Lokasi dengan Named Entity Recognition (NER) ...	20
3.7 Perancangan Pengujian Klasifikasi .....	22
3.7.1 Skenario Pengujian .....	22
3.7.2 Pengelolaan Data Pengujian / Evaluasi.....	23



3.7.3	Accuracy, Recall, dan Precision .....	23
<b>BAB IV</b>	.....	<b>24</b>
<b>IMPLEMENTASI DAN PENGUJIAN</b>	.....	<b>24</b>
4.1.	Implementasi Perangkat Lunak .....	24
4.1.1.	Pengumpulan Data .....	24
4.1.2.	Preprocessing Data.....	26
4.1.3.	Ekstraksi Fitur .....	28
4.1.4.	Klasifikasi .....	31
4.1.5.	Tes Klasifikasi .....	32
4.2.	Hasil Pengujian.....	33
4.2.1.	Hasil Pengujian Model Sintatik .....	33
4.2.2.	Hasil Pengujian Klasifikasi (200 Data Uji).....	33
4.2.3.	Hasil Klasifikasi Fitur Ektraksi Sintatik (200 Data Uji) .....	36
4.3.	Mendapatkan Lokasi .....	44
<b>BAB V</b>	.....	<b>47</b>
<b>PENUTUP</b>	.....	<b>47</b>
5.1.	Kesimpulan.....	47
5.2.	Saran .....	47
<b>DAFTAR PUSTAKA</b>	.....	<b>48</b>



## DAFTAR GAMBAR

Gambar 2.1 Arsitektur Sistem InaNLP .....	12
Gambar 2.2 Ilustrasi dari 10 Fold Cross Validation .....	13
Gambar 3.1 Skema Biosurveillance.....	15
Gambar 3.2 Flowchart Sistem Pre-processing.....	17
Gambar 3.3 Flowchart Sistem Klasifikasi .....	17
Gambar 3.4 Flowchart sistem klasifikasi.....	21
Gambar 4.1 Twitter App Setting.....	24
Gambar 4.2 Advanced Search Twitter.....	25
Gambar 4.3 Hasil Crawling Data Twitter .....	26
Gambar 4.4 Interface Preprocessing Data .....	27
Gambar 4.5 Interface Preprocessing ketika Load Data Twitter.....	27
Gambar 4.6 Interface Hasil Preprocessing Data .....	28
Gambar 4.7 Contoh Pelabelan Data Tweets .....	28
Gambar 4.8 Interface Ekstraksi Fitur Sintaktik .....	29
Gambar 4.9 Data Awal Ekstraksi Fitur Sintaktik .....	29
Gambar 4.10 Contoh kata kunci dari fitur sintaktik .....	30
Gambar 4.11 Data Hasil Ekstraksi Fitur Sintaktik.....	30
Gambar 4.12 Load data train (fitur Sintaktik) pembentukan model.....	31
Gambar 4.13 Notifikasi pembentukan model sukses.....	32
Gambar 4.14 Hasil klasifikasi Fitur Sintaktik.....	32
Gambar 4.15 Cross Validation ekstraksi Fitur Sintaktik .....	33
Gambar 4.16 Hasil summary ekstraksi Fitur Sintaktik.....	34
Gambar 4.17 Interface awal uji klasifikasi manual .....	42
Gambar 4.18 Proses Uji klasifikasi manual.....	43
Gambar 4.19 Hasil perbandingan klasifikasi 200 data test.....	43
Gambar 4.20 Load data awal .....	44
Gambar 4.21 Hasil lokasi yang didapatkan .....	44
Gambar 4.22 Hasil Akurasi Named Entity Recognition.....	45
Gambar 4.23 Hasil Visualisasi Google Maps .....	45
Gambar 4.24 Hasil Data Sorting.....	46

## DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu .....	6
Tabel 2.2 Matriks Konfusi untuk Klasifikasi Dua Kelas .....	14
Tabel 3.1 Daftar Jenis Penyakit .....	16
Tabel 3.2 Tweets Jenis Penyakit .....	16
Tabel 3.3 Data Train .....	18
Tabel 3.4 Daftar Target Kelas .....	18
Tabel 3.5 Daftar Fitur .....	18
Tabel 3.6 Proses ekstraksi fitur dari data train .....	19
Tabel 3.7 Data test .....	20
Tabel 3.8 Proses ekstraksi fitur dari data test .....	20
Tabel 3.9 Hasil perbandingan perhitungan klasifikasi .....	20
Tabel 3.10 Rules Ina-NLP NE-tagger .....	21
Tabel 3.11 Keterangan Rule .....	22
Tabel 3.12 Skenario Pengujian .....	22
Tabel 4.1 Contoh Hasil Data Ekstraksi Fitur .....	28
Tabel 4.2 Confusion matrix dari tes klasifikasi fitur Sintaktik .....	34
Tabel 4.3 Matriks Konfusi kelas InfeksiParasit .....	35
Tabel 4.4 Detail accuracy, precision, recall dari tes klasifikasi fitur Sintaktik .....	35
Tabel 4.5 Hasil Matriks Konfusi kelas "InfeksiVirus" fitur Sintaktik .....	39
Tabel 4.6 Perhitungan Weighted Average Fitur Sintaktik .....	41

## DAFTAR GRAFIK

Grafik 4.1 Hasil true positive semua kelas .....	36
Grafik 4.2 Hasil false positive semua kelas .....	36
Grafik 4.3 Hasil true negative semua kelas.....	37
Grafik 4.4 Hasil false positive semua kelas .....	38
Grafik 4.5 Hasil akurasi (accuracy) tiap kelas .....	38
Grafik 4.6 Hasil presisi (precision) tiap kelas.....	39
Grafik 4.7 Hasil recall tiap kelas.....	40
Grafik 4.8 Hasil weighted average accuracy, precision, recall seluruh kelas.....	40



## DAFTAR PUSTAKA

- [1] NACCHO, "National Association of Country & City Health Officials," 2012. [Online]. Available: <http://archived.naccho.org/topics/emergency/biosurveillance/>.
- [2] F. -C. Tsui, "Technical Description of RODS: A Real-time Public Health Surveillance System," 2003.
- [3] D. Gaffney and C. Puschmann, "Data collection on Twitter," pp. 55-68, 2013.
- [4] S. Kumar, F. Morstatter and H. Liu, "Twitter data analytics. Springer Briefs in Computer," 2015.
- [5] S. Asur and B. A. Huberman, "Predicting the Future with Social Media," 2010.
- [6] F. M. F. Wong, S. Sen and M. Chiang, "Why watching movie tweets won't tell the whole story/," 2012.
- [7] D. Gayo-Avello, "'I wanted to predict elections with Twitter and all I got was this lousy paper': A balanced survey on election prediction using Twitter data.," 2012.
- [8] H. Zhang, "The Optimality of Naive Bayes," 2004.
- [9] R. E. Walpole and R. H. Myers, "Ilmu peluang dan Statistika untuk Insinyur dan Ilmuan Bandung:ITB," 1995.
- [10] Syarli and A. A. Muin, "Metode Naive Bayes untuk Prediksi Kelulusan(Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi)," 2016.
- [11] W. House, "National Strategy for Biosurveillance," 2012.
- [12] R. Feldman and J. Sanger, The Text Mining Handbook, Cambridge University Press, 2007.
- [13] J. Han and M. Kember, "Data Mining : Concepts and Techniques," 2006.
- [14] B. Loni, "Enhanced Question Classification with Optimal Combination of Features: A New Approach on Automated Question Answering Systems," *Pattern Recognit*, 2012.

- [15] L. Hadjaratie, "Jaringan Saraf Tiruan untuk Prediksi Tingkat Kelulusan Mahasiswa Diploma Program Studi Manajemen Informatika Universitas Negeri Gorontalo," 2011.
- [16] A. Indranandita, B. Susanto and A. R. C., "Sistem Klasifikasi dan Pencarian Jurnal dengan Menggunakan Metode Naive Bayes dan Vector Space Model.," 2008.
- [17] N. Chinchor, E. Brown, L. Ferro and P. Robinson, Named Entity Recognition Task Definition, 1999.
- [18] A. Purwarianti, A. F. Wicaksono, A. Andhika, I. Afif and F. Ferdian, "InaNLP: Indonesia Natural Language Processing Toolkit (Case study: Complaint Tweet Classification)," 2016.
- [19] E. Prasetyo, Data Mining-Konsep dan Aplikasi Menggunakan MATLAB, Yogyakarta: ANDI, 2012.
- [20] A. Wibowo, "Bina Nusantara," 24 11 2017. [Online]. Available: <https://mti.binus.ac.id/2017/11/24/10-fold-cross-validation/>.